Polarization codes and the rate of polarization

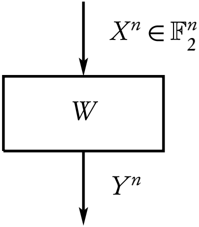
Erdal Arıkan, Emre Telatar

Bilkent U., EPFL

December 3, 2009

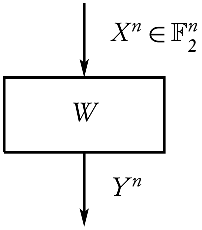
Channel Polarization

- Given a binary input DMC W,



Channel Polarization

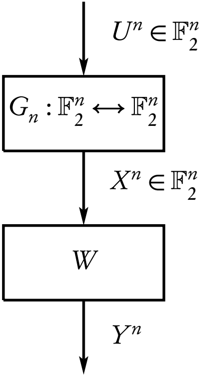
- Given a binary input DMC W,
- i.i.d. uniformly distributed inputs $(X_1, \ldots, X_n) \in \{0, 1\}^n$,



ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

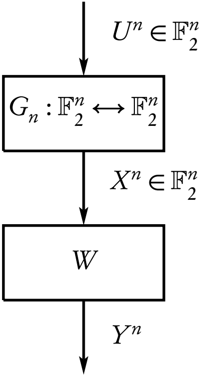
Channel Polarization

- Given a binary input DMC W,
- i.i.d. uniformly distributed inputs $(X_1, \ldots, X_n) \in \{0, 1\}^n$,
- in one-to-one correspondence with binary 'data' $(U_1, \ldots, U_n) \in \{0, 1\}^n$.



Channel Polarization

- Given a binary input DMC W,
- i.i.d. uniformly distributed inputs $(X_1, \ldots, X_n) \in \{0,1\}^n$,
- in one-to-one correspondence with binary 'data' $(U_1, \ldots, U_n) \in \{0,1\}^n$.
- Observe that U_i are i.i.d., uniform on $\{0,1\}$.



Channel Polarization

$$I(U^n; Y^n) =$$

Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n)$$

Channel Polarization

$$\begin{aligned} I(U^n;Y^n) &= I(X^n;Y^n) \\ &= \sum_i I(X_i;Y_i) \end{aligned}$$

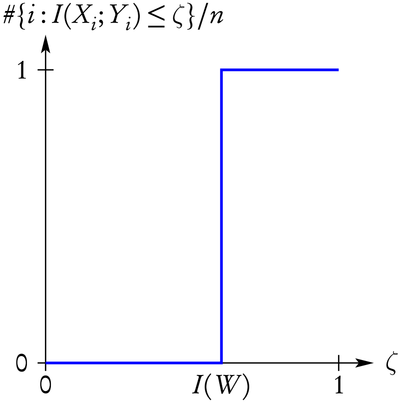
Channel Polarization

$$
\begin{aligned}
I(U^n;Y^n) &= I(X^n;Y^n) \\
&= \sum_i I(X_i;Y_i) \\
&= \sum_i I(W)
\end{aligned}
$$

Notation: $I(P)$ denotes the mutual information between the input and output of a channel P when input is uniformly distributed.

Channel Polarization

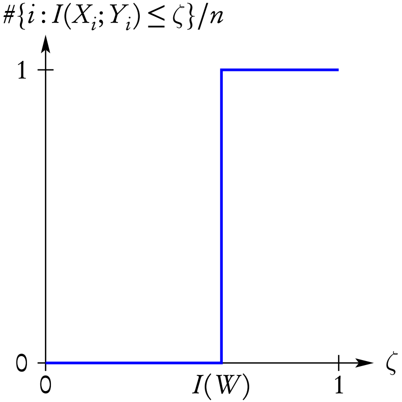
$$
\begin{aligned}
I(U^n;Y^n) &= I(X^n;Y^n) \\
&= \sum_i I(X_i;Y_i) \\
&= \sum_i I(W)
\end{aligned}
$$



Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n) = \sum_i I(X_i;Y_i) = \sum_i I(W)$$

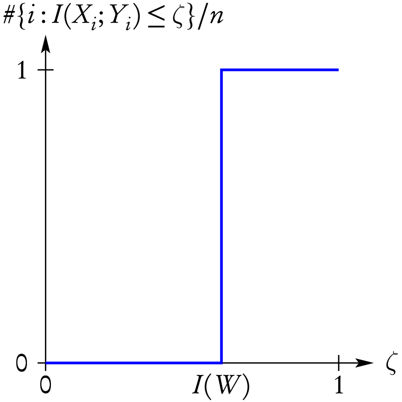
$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1})$$



Channel Polarization

$$
\begin{aligned}
I(U^n;Y^n) &= I(X^n;Y^n) \\
&= \sum_i I(X_i;Y_i) \\
&= \sum_i I(W)
\end{aligned}
$$

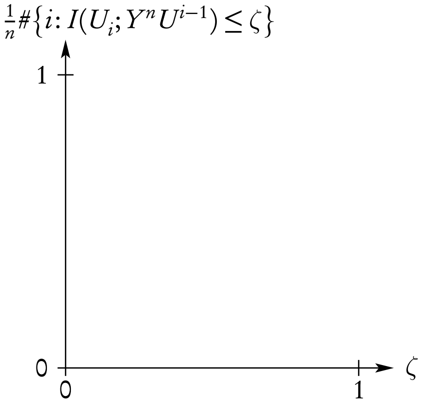
$$
\begin{aligned}
I(U^n;Y^n) &= \sum_i I(U_i;Y^n|U^{i-1}) \\
&= \sum_i I(U_i;Y^n U^{i-1})
\end{aligned}
$$



Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n)$$
$$= \sum_i I(X_i;Y_i)$$
$$= \sum_i I(W)$$

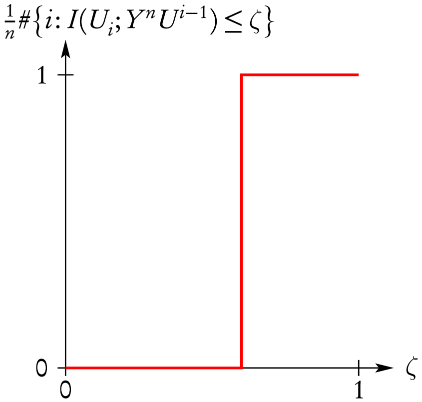
$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1})$$
$$= \sum_i I(U_i;Y^n U^{i-1})$$



Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n) = \sum_i I(X_i;Y_i) = \sum_i I(W)$$

$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1}) = \sum_i I(U_i;Y^nU^{i-1})$$

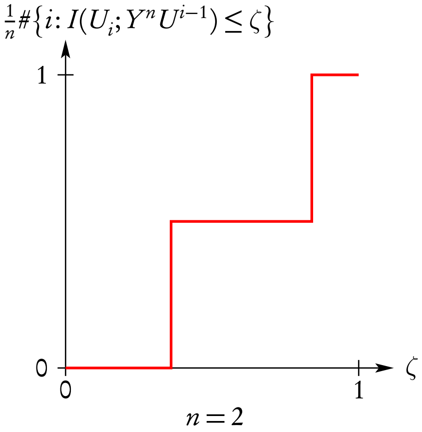


$n = 1$

Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n)$$
$$= \sum_i I(X_i;Y_i)$$
$$= \sum_i I(W)$$

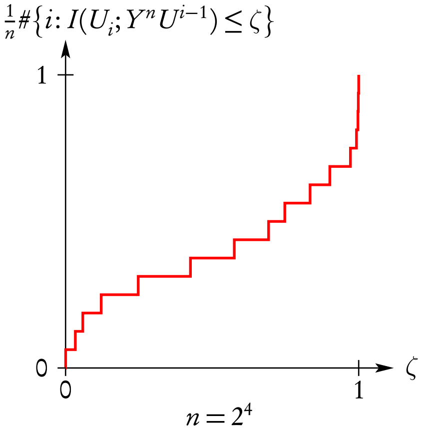
$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1})$$
$$= \sum_i I(U_i;Y^nU^{i-1})$$



Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n) = \sum_i I(X_i;Y_i) = \sum_i I(W)$$

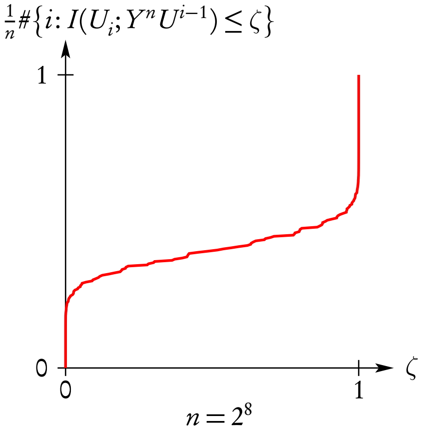
$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1}) = \sum_i I(U_i;Y^nU^{i-1})$$



Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n)$$
$$= \sum_i I(X_i;Y_i)$$
$$= \sum_i I(W)$$

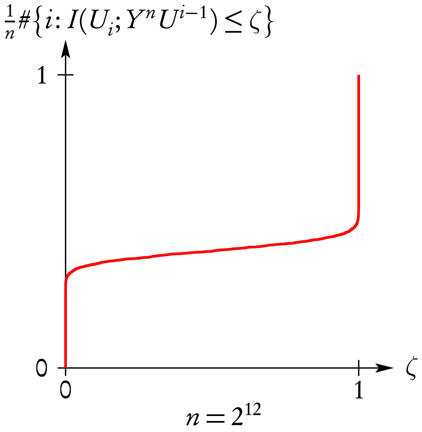
$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1})$$
$$= \sum_i I(U_i;Y^nU^{i-1})$$



Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n) = \sum_i I(X_i;Y_i) = \sum_i I(W)$$

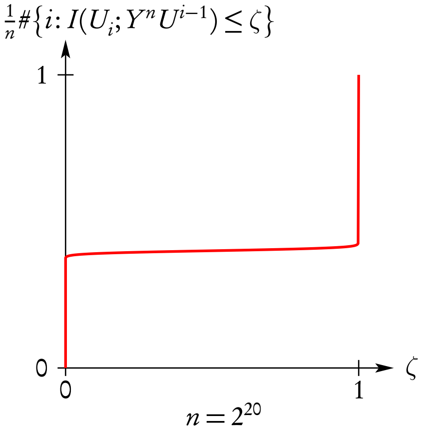
$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1}) = \sum_i I(U_i;Y^nU^{i-1})$$



Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n)$$
$$= \sum_i I(X_i;Y_i)$$
$$= \sum_i I(W)$$

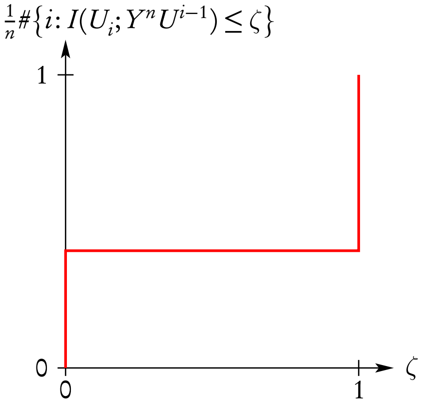
$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1})$$
$$= \sum_i I(U_i;Y^nU^{i-1})$$



Channel Polarization

$$I(U^n;Y^n) = I(X^n;Y^n)$$
$$= \sum_i I(X_i;Y_i)$$
$$= \sum_i I(W)$$

$$I(U^n;Y^n) = \sum_i I(U_i;Y^n|U^{i-1})$$
$$= \sum_i I(U_i;Y^nU^{i-1})$$



Channel Polarization

We say channel polarization takes place if it is the case that almost all of the numbers $I(U_i; Y^n U^{i-1})$ are near the extremal values,

$$\frac{1}{n}\#\{i : I(U_i; Y^n U^{i-1}) \in (\epsilon, 1-\epsilon)\} \to 0.$$

Channel Polarization

We say channel polarization takes place if it is the case that almost all of the numbers $I(U_i; Y^n U^{i-1})$ are near the extremal values,

$$\frac{1}{n}\#\{i: I(U_i; Y^n U^{i-1}) \in (\epsilon, 1-\epsilon)\} \to 0.$$

Equivalently, if

$$\frac{1}{n}\#\{i: I(U_i; Y^n U^{i-1}) \approx 1\} \to I(W)$$

and

$$\frac{1}{n}\#\{i: I(U_i; Y^n U^{i-1}) \approx 0\} \to 1 - I(W).$$

Channel Polarization

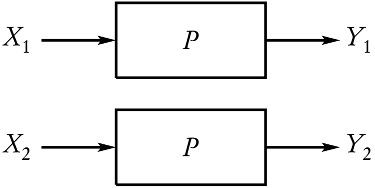
If polarization takes place and we wish to communicate at rate R:

- Pick n, and $k = nR$ good indices i such that $I(U_i; Y^n U^{i-1})$ is high,
- let the transmitter set U_i to be uncoded binary data for good indices, and set U_i to random but publicly known values for the rest,
- let the receiver decode the U_i successively: U_1 from Y^n; U_i from $Y^n \hat{U}^{i-1}$.
- One would expect this scheme to do well as long as there are k good indices, i.e., if $R < I(W)$.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

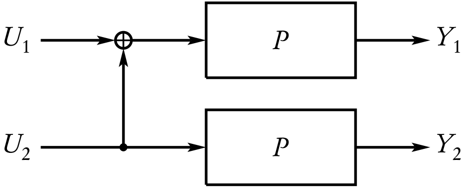
Polarization – HowTo

Given two copies of a binary input channel $P\colon \mathbb{F}_2 \to \mathscr{Y}$



Polarization – HowTo

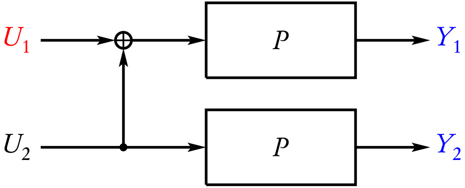
Given two copies of a binary input channel $P\colon \mathbb{F}_2 \to \mathscr{Y}$



consider the transformation above to generate two channels

Polarization – HowTo

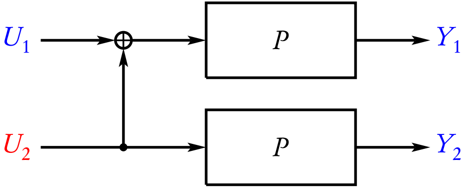
Given two copies of a binary input channel $P\colon \mathbb{F}_2 \to \mathcal{Y}$



consider the transformation above to generate two channels
$P^- : F_2 \to \mathcal{Y}^2$

Polarization – HowTo

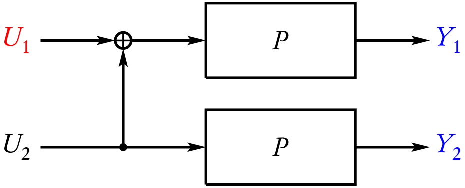
Given two copies of a binary input channel $P\colon \mathbb{F}_2 \to \mathscr{Y}$



consider the transformation above to generate two channels $P^{-}: F_2 \to \mathscr{Y}^2$ and $P^{+}: F_2 \to \mathscr{Y}^2 \times F_2$

Polarization – HowTo

Given two copies of a binary input channel $P\colon \mathbb{F}_2 \to \mathcal{Y}$

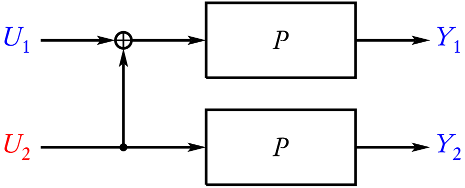


consider the transformation above to generate two channels $P^{-}: F_2 \to \mathcal{Y}^2$ and $P^{+}: F_2 \to \mathcal{Y}^2 \times F_2$ with

$$P^{-}(y_1 y_2 | u_1) = \sum_{u_2} \tfrac{1}{2} P(y_1 | u_1 + u_2) P(y_2 | u_2)$$

Polarization – HowTo

Given two copies of a binary input channel $P\colon \mathbb{F}_2 \to \mathscr{Y}$



consider the transformation above to generate two channels $P^- : F_2 \to \mathscr{Y}^2$ and $P^+ : F_2 \to \mathscr{Y}^2 \times F_2$ with

$$P^-(y_1 y_2 | u_1) = \sum_{u_2} \tfrac{1}{2} P(y_1 | u_1 + u_2) P(y_2 | u_2)$$

$$P^+(y_1 y_2 u_1 | u_2) = \tfrac{1}{2} P(y_1 | u_1 + u_2) P(y_2 | u_2).$$

Polarization – HowTo

- Observe that
$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

Polarization – HowTo

- Observe that

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

- With independent, uniform U_1, U_2,

$$I(P^-) = I(U_1; Y_1 Y_2),$$
$$I(P^+) = I(U_2; Y_1 Y_2 U_1).$$

Polarization – HowTo

- Observe that

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

- With independent, uniform U_1, U_2,

$$I(P^-) = I(U_1; Y_1 Y_2),$$
$$I(P^+) = I(U_2; Y_1 Y_2 U_1).$$

- Thus,

$$I(P^-) + I(P^+) = I(U_1 U_2; Y_1 Y_2) = 2I(P),$$

Polarization – HowTo

- Observe that

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

- With independent, uniform U_1, U_2,

$$I(P^-) = I(U_1; Y_1 Y_2),$$
$$I(P^+) = I(U_2; Y_1 Y_2 U_1).$$

- Thus,

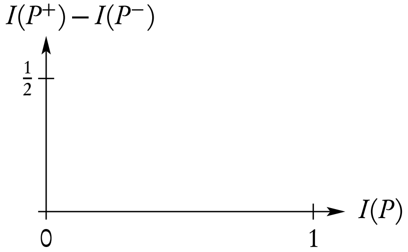
$$I(P^-) + I(P^+) = I(U_1 U_2; Y_1 Y_2) = 2I(P),$$

- and $I(P^-) \leq I(P) \leq I(P^+)$.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

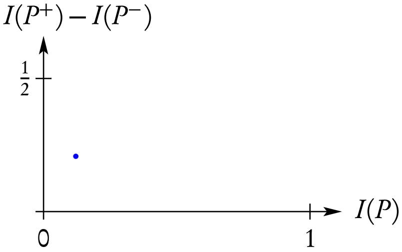
Polarization – Howto

How far apart are $I(P^-)$ and $I(P^+)$?



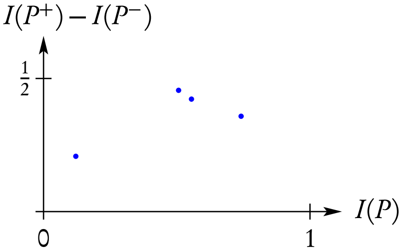
Polarization – Howto

How far apart are $I(P^-)$ and $I(P^+)$?



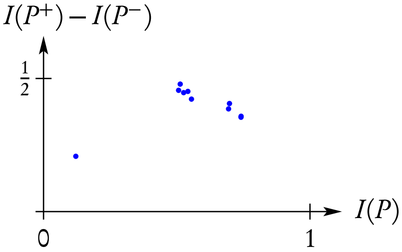
Polarization – Howto

How far apart are $I(P^-)$ and $I(P^+)$?



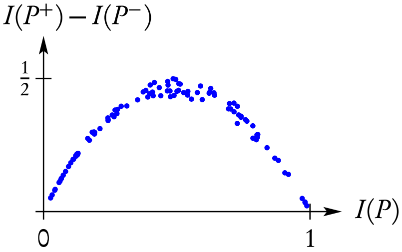
Polarization – Howto

How far apart are $I(P^-)$ and $I(P^+)$?



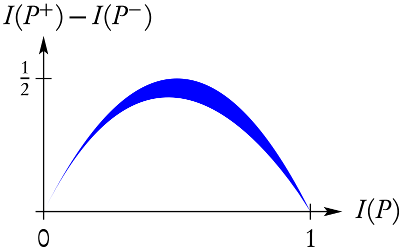
Polarization – Howto

How far apart are $I(P^-)$ and $I(P^+)$?



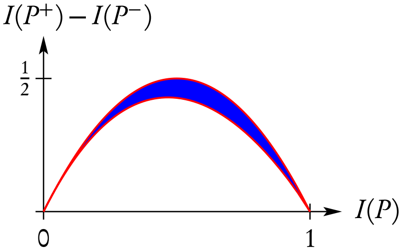
Polarization – Howto

How far apart are $I(P^-)$ and $I(P^+)$?



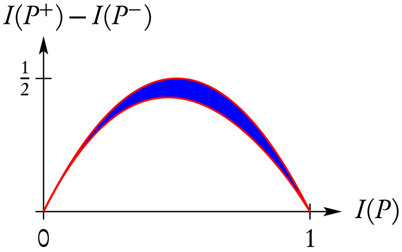
Polarization – Howto

How far apart are $I(P^-)$ and $I(P^+)$?



Polarization – Howto

How far apart are $I(P^-)$ and $I(P^+)$?



$I(P^+) - I(P^-) < \epsilon$ implies that $I(P) \notin (\delta, 1-\delta)$.

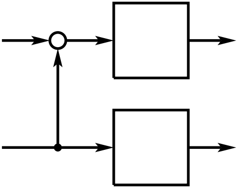
Polarization – HowTo

What we can do once, we can do many times: Given W,

Polarization – HowTo

What we can do once, we can do many times: Given W,

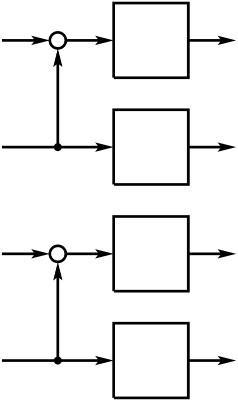
- Duplicate W and obtain W^- and W^+.



Polarization – HowTo

What we can do once, we can do many times: Given W,

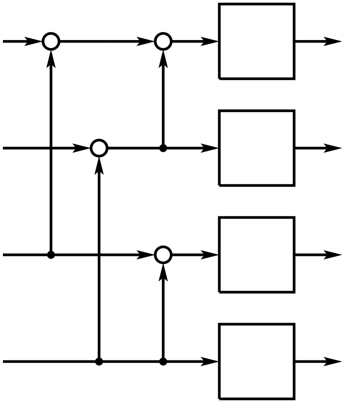
- Duplicate W and obtain W^- and W^+.
- Duplicate W^- (W^+),



Polarization – HowTo

What we can do once, we can do many times: Given W,

- Duplicate W and obtain W^{-} and W^{+}.
- Duplicate W^{-} (W^{+}),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).

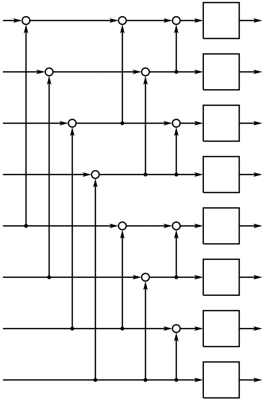


ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Polarization – HowTo

What we can do once, we can do many times: Given W,

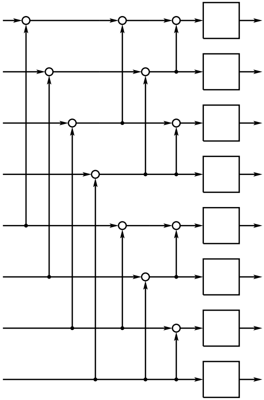
- Duplicate W and obtain W^{-} and W^{+}.
- Duplicate W^{-} (W^{+}),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).
- Duplicate W^{--} (W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).



Polarization – HowTo

What we can do once, we can do many times: Given W,

- Duplicate W and obtain W^- and W^+.
- Duplicate W^- (W^+),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).
- Duplicate W^{--} (W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).
- …



Polarization – HowTo

- We now have a one-to-one transformation G_n between $U_1, \ldots, U_n$ and $X_1, \ldots, X_n$ for $n = 2^\ell$.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Polarization – HowTo

- We now have a one-to-one transformation G_n between $U_1, \ldots, U_n$ and $X_1, \ldots, X_n$ for $n = 2^\ell$.
- Furthermore, the $n = 2^\ell$ quantities

$$I(W^{b_1 \ldots, b_\ell}), \quad b_j \in \{+, -\}$$

Polarization – HowTo

- We now have a one-to-one transformation G_n between $U_1, \ldots, U_n$ and $X_1, \ldots, X_n$ for $n = 2^\ell$.
- Furthermore, the $n = 2^\ell$ quantities

$$I(W^{b_1 \ldots, b_\ell}), \quad b_j \in \{+, -\}$$

- are exactly the n quantities

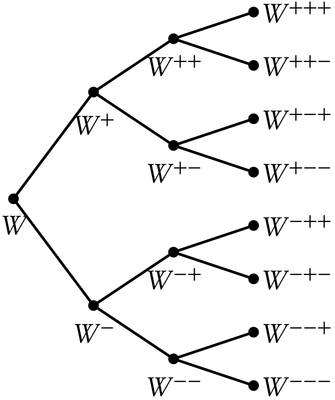
$$I(U_i; Y^n U^{i-1}), \quad i = 1, \ldots, n$$

whose empirical distribution we seek.

Polarization – HowTo

This suggests the following:

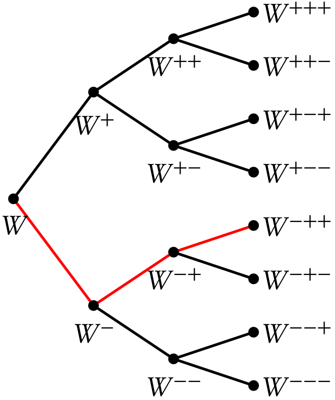
- Let $B_1, B_2, \ldots$ be i.i.d., equally likely to be $\{+, -\}$, $W_0 = W$, $W_\ell = W_{\ell-1}^{B_\ell}$.



Polarization – HowTo

This suggests the following:

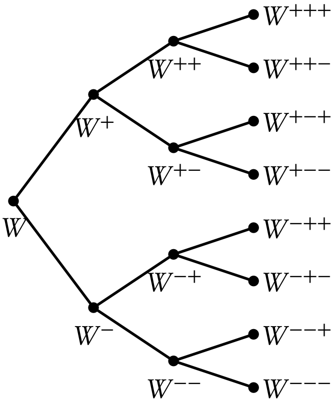
- Let $B_1, B_2, \ldots$ be i.i.d., equally likely to be $\{+, -\}$, $W_0 = W$, $W_\ell = W_{\ell-1}^{B_\ell}$.



Polarization – HowTo

This suggests the following:

- Let $B_1, B_2, \ldots$ be i.i.d., equally likely to be $\{+,-\}$, $W_0 = W$, $W_\ell = W_{\ell-1}^{B_\ell}$.
- W_ℓ is uniformly distributed among $\left\{W^{-\cdots-}, \ldots, W^{+\cdots+}\right\}$,

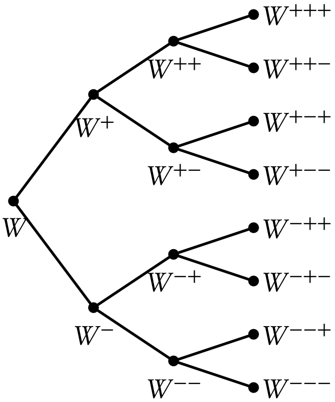


Polarization – HowTo

This suggests the following:

- Let $B_1, B_2, \ldots$ be i.i.d., equally likely to be $\{+,-\}$, $W_0 = W$, $W_\ell = W_{\ell-1}^{B_\ell}$.
- W_ℓ is uniformly distributed among $\{W^{-\cdots-}, \ldots, W^{+\cdots+}\}$,
- $I_\ell = I(W_\ell)$ is distributed as

$$\Pr(I_\ell \in E) = \frac{1}{n} \#\{i : I(U_i; Y^n U^{i-1}) \in E\}.$$



Polarization – HowTo

Properties of the process I_ℓ:

- $I_0 = I(W)$ is a constant.

Polarization – HowTo

Properties of the process I_ℓ:

- $I_0 = I(W)$ is a constant.
- I_ℓ is lies in $[0, 1]$, so is bounded.

Polarization – HowTo

Properties of the process I_ℓ:

- $I_0 = I(W)$ is a constant.
- I_ℓ is lies in $[0,1]$, so is bounded.
- Conditional on $B_1,\dots,B_\ell$, we know W_ℓ, and $I_{\ell+1}$ is equally likely to be $I(W_\ell^-)$ and $I(W_\ell^+)$,

Polarization – HowTo

Properties of the process I_ℓ:

- $I_0 = I(W)$ is a constant.
- I_ℓ is lies in $[0,1]$, so is bounded.
- Conditional on $B_1,\dots,B_\ell$, we know W_ℓ, and $I_{\ell+1}$ is equally likely to be $I(W_\ell^-)$ and $I(W_\ell^+)$,
- so,

$$E[I_{\ell+1}|B_1,\dots,B_\ell] = \frac{1}{2}[I(W_\ell^-)+I(W_\ell^+)] = I(W_\ell) = I_\ell$$

and we see that $\{I_\ell\}$ is a martingale.

Polarization – HowTo

Properties of the process I_ℓ:

- $I_0 = I(W)$ is a constant.
- I_ℓ is lies in $[0,1]$, so is bounded.
- Conditional on $B_1,\dots,B_\ell$, we know W_ℓ, and $I_{\ell+1}$ is equally likely to be $I(W_\ell^-)$ and $I(W_\ell^+)$,
- so,

$$E[I_{\ell+1}|B_1,\dots,B_\ell] = \tfrac{1}{2}[I(W_\ell^-)+I(W_\ell^+)] = I(W_\ell) = I_\ell$$

  and we see that $\{I_\ell\}$ is a martingale.
- Recall that what we are trying to show is that I_ℓ converges weakly to a $\{0,1\}$-valued random variable.

Polarization – HowTo

- Bounded martingales converge almost surely. Thus $I_\infty := \lim_{\ell \to \infty} I_\ell$ exists with probability one. So we know that the empirical distribution we seek exists.

Polarization – HowTo

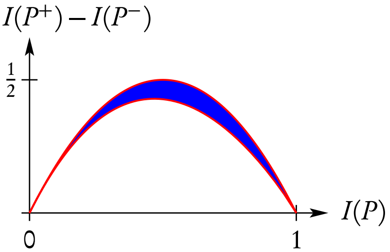
- Bounded martingales converge almost surely. Thus $I_\infty := \lim_{\ell \to \infty} I_\ell$ exists with probability one. So we know that the empirical distribution we seek exists.
- Almost sure convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$ a.s.

Polarization – HowTo

- Bounded martingales converge almost surely. Thus $I_\infty := \lim_{\ell\to\infty} I_\ell$ exists with probability one. So we know that the empirical distribution we seek exists.
- Almost sure convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$ a.s.
- $|I_{\ell+1} - I_\ell| = \frac{1}{2}[I(P^+) - I(P^-)]$.

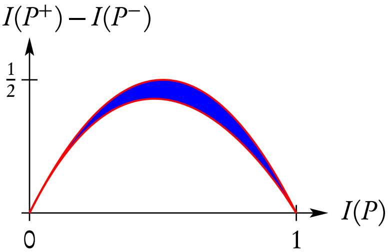
Polarization – HowTo

- Bounded martingales converge almost surely. Thus $I_\infty := \lim_{\ell\to\infty} I_\ell$ exists with probability one. So we know that the empirical distribution we seek exists.
- Almost sure convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$ a.s.
- $|I_{\ell+1} - I_\ell| = \frac{1}{2}[I(P^+) - I(P^-)]$.
- But $I(P^+) - I(P^-) < \epsilon$ implies $I(P) \notin (\delta, 1-\delta)$.



Polarization – HowTo

- Bounded martingales converge almost surely. Thus $I_\infty := \lim_{\ell \to \infty} I_\ell$ exists with probability one. So we know that the empirical distribution we seek exists.
- Almost sure convergence of $\{I_\ell\}$ implies $|I_{\ell+1} - I_\ell| \to 0$ a.s.
- $|I_{\ell+1} - I_\ell| = \frac{1}{2}[I(P^+) - I(P^-)]$.
- But $I(P^+) - I(P^-) < \epsilon$ implies $I(P) \notin (\delta, 1-\delta)$.
- Thus I_∞ is $\{0, 1\}$ valued!



Polarization Rate

- We have seen that polarization takes place.

Polarization Rate

- We have seen that polarization takes place.
- But how fast? Fast enough to arrest error propagation?

Polarization Rate

- We have seen that polarization takes place.
- But how fast? Fast enough to arrest error propagation?
- Introduce an auxiliary quantity

$$Z(P) = \sum_{y} \sqrt{P(y|0)P(y|1)}$$

as a companion to $I(P)$. Note that this is the Bhattacharyya upper bound on probability of error for uncoded transmission over P.

Polarization Rate

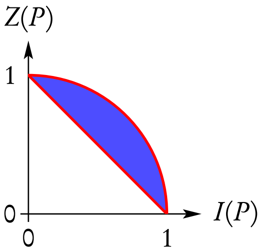
Properties of $Z(P)$:

- $Z(P) \in [0, 1]$.

Polarization Rate

Properties of $Z(P)$:

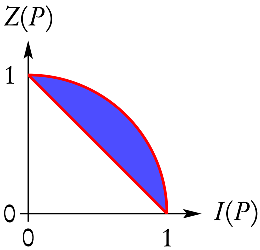
- $Z(P) \in [0, 1]$.
- $Z(P) \approx 0$ iff $I(P) \approx 1$.



Polarization Rate

Properties of $Z(P)$:

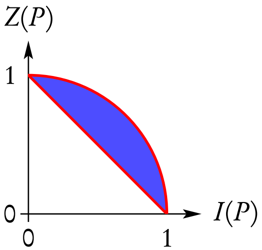
- $Z(P) \in [0,1]$.
- $Z(P) \approx 0$ iff $I(P) \approx 1$.
- $Z(P) \approx 1$ iff $I(P) \approx 0$.



Polarization Rate

Properties of $Z(P)$:

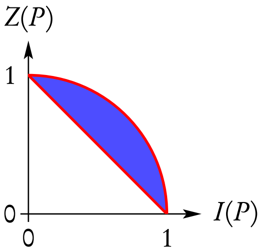
- $Z(P) \in [0, 1]$.
- $Z(P) \approx 0$ iff $I(P) \approx 1$.
- $Z(P) \approx 1$ iff $I(P) \approx 0$.
- $Z(P^+) = Z(P)^2$.



Polarization Rate

Properties of $Z(P)$:

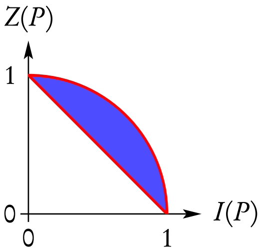
- $Z(P) \in [0,1]$.
- $Z(P) \approx 0$ iff $I(P) \approx 1$.
- $Z(P) \approx 1$ iff $I(P) \approx 0$.
- $Z(P^+) = Z(P)^2$.
- $Z(P^-) \leq 2Z(P)$.



Polarization Rate

Properties of $Z(P)$:

- $Z(P) \in [0, 1]$.
- $Z(P) \approx 0$ iff $I(P) \approx 1$.
- $Z(P) \approx 1$ iff $I(P) \approx 0$.
- $Z(P^+) = Z(P)^2$.
- $Z(P^-) \leq 2Z(P)$.



Since $Z(P)$ an upper bound on probability of error for uncoded transmission over P, we can choose the good indices on the basis of $Z(P)$. The sum of the Z's of the chosen channels will upper bound the block error probability. This suggests studying the polarization rate of Z.

Polarization Rate

Given a binary input channel W,

Polarization Rate

Given a binary input channel W,

- just as for I_ℓ, define $Z_\ell = Z(W^{B_1,\dots,B_\ell})$.

Polarization Rate

Given a binary input channel W,

- just as for I_ℓ, define $Z_\ell = Z(W^{B_1,\dots,B_\ell})$.
- We know that $\Pr(Z_\ell \to 0) = I(W)$.

Polarization Rate

Given a binary input channel W,

- just as for I_ℓ, define $Z_\ell = Z(W^{B_1,\dots,B_\ell})$.
- We know that $\Pr(Z_\ell \to 0) = I(W)$.
- It turns out that when $Z_\ell \to 0$ it does so fast:

Theorem

For any $\beta < 1/2$, $\lim_{\ell\to\infty} \Pr(Z_\ell < 2^{-2^{\beta\ell}}) = I(W)$.

Polarization Rate

Given a binary input channel W,

- just as for I_ℓ, define $Z_\ell = Z(W^{B_1,\dots,B_\ell})$.
- We know that $\Pr(Z_\ell \to 0) = I(W)$.
- It turns out that when $Z_\ell \to 0$ it does so fast:

Theorem

For any $\beta < 1/2$, $\lim_{\ell\to\infty} \Pr(Z_\ell < 2^{-2^{\beta\ell}}) = I(W)$.

- This means that for any $\beta < 1/2$, as long as $R < I(W)$ the error probability of polarization codes decays to 0 faster than 2^{-n^β}.

Polarization Rate

Proof idea

- The subset of the probability space where $Z_\ell(\omega) \to 0$ has probability $I(W)$.

Polarization Rate

Proof idea

- The subset of the probability space where $Z_\ell(\omega) \to 0$ has probability $I(W)$.
- For any $\epsilon > 0$ there is an m and a subset of this set with probability at least $I(W) - \epsilon$ with $Z_\ell(\omega) < 1/8$ whenever $\ell > m$.

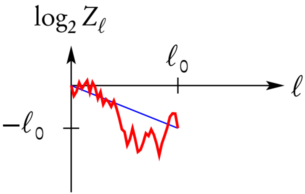
Polarization Rate

Proof idea

- The subset of the probability space where $Z_\ell(\omega) \to 0$ has probability $I(W)$.
- For any $\epsilon > 0$ there is an m and a subset of this set with probability at least $I(W) - \epsilon$ with $Z_\ell(\omega) < 1/8$ whenever $\ell > m$.
- Intersect this set with high probability event: that '$B_i = +$' and '$B_i = -$' occur with almost equal frequency for $i \le \ell$. It is then easy to see that $\lim_{\ell\to\infty} P(\log_2 Z_\ell \le -\ell) = I(W)$.

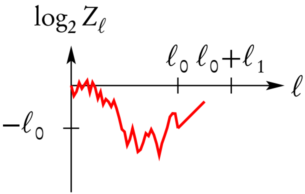
Polarization Rate

Proof idea – cont.



Polarization Rate

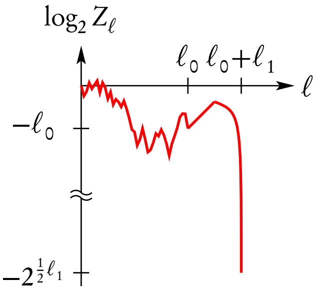
Proof idea – cont.



$$\ell_1 = \ell_0$$

Polarization Rate

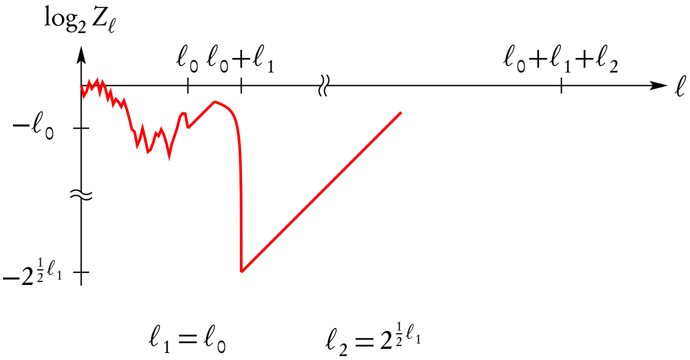
Proof idea – cont.



$$\ell_1 = \ell_0$$

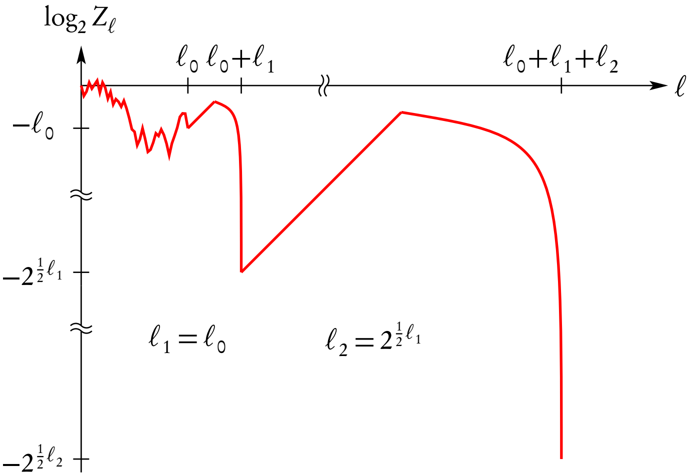
Polarization Rate

Proof idea – cont.



Polarization Rate

Proof idea – cont.



Remarks

With the particular one-to-one mapping described here and with successive cancellation decoding

Remarks

With the particular one-to-one mapping described here and with successive cancellation decoding

- polarization codes are '$I(W)$ achieving',

Remarks

With the particular one-to-one mapping described here and with successive cancellation decoding

- polarization codes are '$I(W)$ achieving',
- encoding complexity is $n \log n$,

Remarks

With the particular one-to-one mapping described here and with successive cancellation decoding

- polarization codes are '$I(W)$ achieving',
- encoding complexity is $n \log n$,
- decoding complexity is $n \log n$,

Remarks

With the particular one-to-one mapping described here and with successive cancellation decoding

- polarization codes are '$I(W)$ achieving',
- encoding complexity is $n \log n$,
- decoding complexity is $n \log n$,
- probability of error decays like $2^{-\sqrt{n}}$.

Remarks

With the particular one-to-one mapping described here and with successive cancellation decoding

- polarization codes are '$I(W)$ achieving',
- encoding complexity is $n \log n$,
- decoding complexity is $n \log n$,
- probability of error decays like $2^{-\sqrt{n}}$.
- However, for small ℓ their performance is surpassed by the usual suspects.

Remarks

Moreover:

- For symmetric channels the ‘dummy’ symbols can be chosen to be constant, yielding a deterministic construction,

Remarks

Moreover:

- For symmetric channels the ‘dummy’ symbols can be chosen to be constant, yielding a deterministic construction,
- the error probability guarantees are not based on simulation.

Remarks

Moreover:

- For symmetric channels the 'dummy' symbols can be chosen to be constant, yielding a deterministic construction,
- the error probability guarantees are not based on simulation.
- They can be generalized to channels with arbitrary discrete input alphabets: a similar 'two-by-two' construction, with same complexity and error probability bounds.

Remarks

Moreover:

- For symmetric channels the 'dummy' symbols can be chosen to be constant, yielding a deterministic construction,
- the error probability guarantees are not based on simulation.
- They can be generalized to channels with arbitrary discrete input alphabets: a similar 'two-by-two' construction, with same complexity and error probability bounds.
- With 'k-by-k' constructions the error probability can be made to decay almost exponentially (Korada, Şaşoğlu, Urbanke, 2009)

Remarks

Moreover:

- For symmetric channels the 'dummy' symbols can be chosen to be constant, yielding a deterministic construction,
- the error probability guarantees are not based on simulation.
- They can be generalized to channels with arbitrary discrete input alphabets: a similar 'two-by-two' construction, with same complexity and error probability bounds.
- With 'k-by-k' constructions the error probability can be made to decay almost exponentially (Korada, Şaşoğlu, Urbanke, 2009)
- Can be adapted to the multiple-access channel.

Remarks

Moreover:

- For symmetric channels the 'dummy' symbols can be chosen to be constant, yielding a deterministic construction,
- the error probability guarantees are not based on simulation.
- They can be generalized to channels with arbitrary discrete input alphabets: a similar 'two-by-two' construction, with same complexity and error probability bounds.
- With 'k-by-k' constructions the error probability can be made to decay almost exponentially (Korada, Şaşoğlu, Urbanke, 2009)
- Can be adapted to the multiple-access channel.
- Dual constructions give quantizers.

Open Questions

Many open questions. A few:

Open Questions

Many open questions. A few:

- Improve the error probability estimate to include rate dependence.

Open Questions

Many open questions. A few:

- Improve the error probability estimate to include rate dependence.
- Polarization appears to be a general phenomenon, how general?

Open Questions

Many open questions. A few:

- Improve the error probability estimate to include rate dependence.
- Polarization appears to be a general phenomenon, how general?
- Successive cancellation is too naive, different decoding algorithms should do better.

Open Questions

Many open questions. A few:

- Improve the error probability estimate to include rate dependence.
- Polarization appears to be a general phenomenon, how general?
- Successive cancellation is too naive, different decoding algorithms should do better.
- Numerical evidence suggests scaling laws for the distribution of I_ℓ and Z_ℓ. Would be nice to have them.

Open Questions

Many open questions. A few:

- Improve the error probability estimate to include rate dependence.
- Polarization appears to be a general phenomenon, how general?
- Successive cancellation is too naive, different decoding algorithms should do better.
- Numerical evidence suggests scaling laws for the distribution of I_ℓ and Z_ℓ. Would be nice to have them.
- Multi-user applications.